3.3 Classification

Labelled features of trained data (HSV images) are then stored in the database. Testing is performed by entering unlabelled images of *ambon lumut, kepok* and *raja* bananas, convert them into HSV, having their features extracted and then performed classification process. The steps for class determination are shown in Figure 2, while the Data Flow Diagrams (DFD) consisting of Context DFD, Overview DFD and Primitive DFD are shown in Figure 3 to 5.

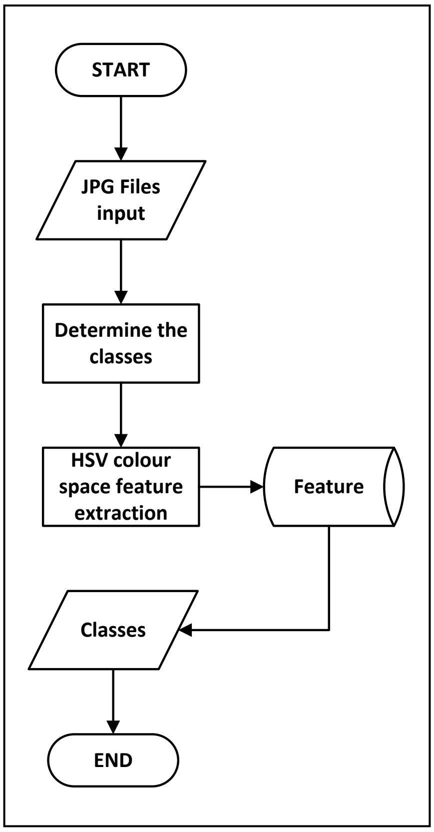

Figure 2: Class determination

Context Level 0 DFD describes the overview of the system which has a single external entity (USER). This entity provides input for the system in form of training and testing data in JPG image format. The system will then give the output of classes and the classification results, as shown in Figure 3.

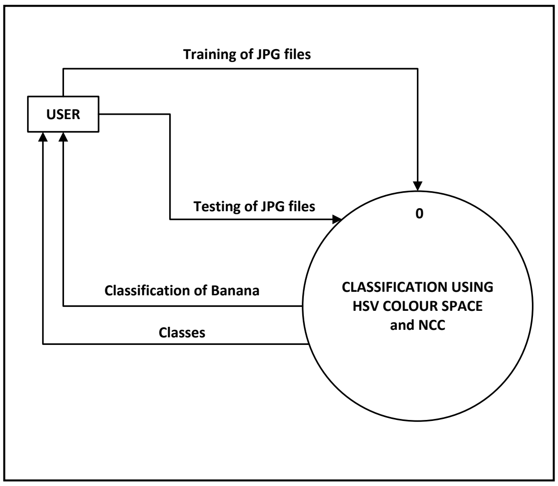

Figure 3: Context DFD of Classification Using HSV Colour Space

Breakdown of the context DFD is the Overview Diagram which describes processes exist within the system. This diagram is also called Zero Diagram or Level-0 which as shown in Figure 4, it has two processes. The first process 1.0 is the process which accepts input in the form of JPG images for training, extract its HSV colour features and store them into the Feature data store. This process also provides output in form of classes returned to external entity USER. The other process is 2.0 which is the classifier. When it is accepting input from 1.0 as a testing data, it reads the stored features from Feature data store and performs classification. The resulting classification is then returned as output to external entity USER.

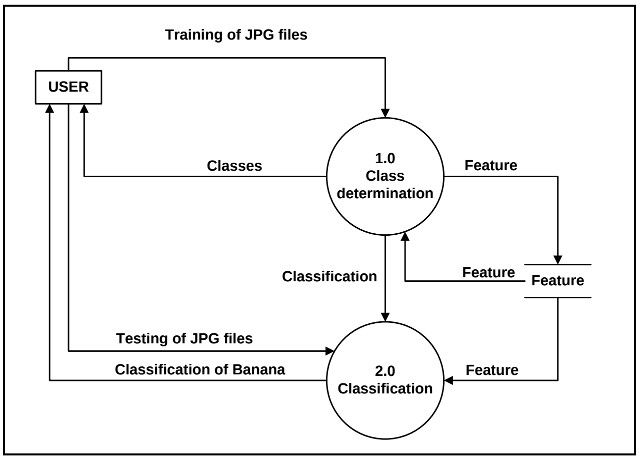

Figure 4: Overview DFD of Classification Using HSV Colour Space

The two processes can further be described using two separate diagrams of Primitive DFD or Level-1 DFD which each showed the detailed processes of Process 1 and Process 2. The breakdown of Process 1.0 are Process 1.1 p which accepts the JPG files input, extract its features and store them to Feature data store; and Process 1.2 p which reads the Feature data store and performing class determination, giving the output of classes to the external entity USER, as illustrated in Figure 5.

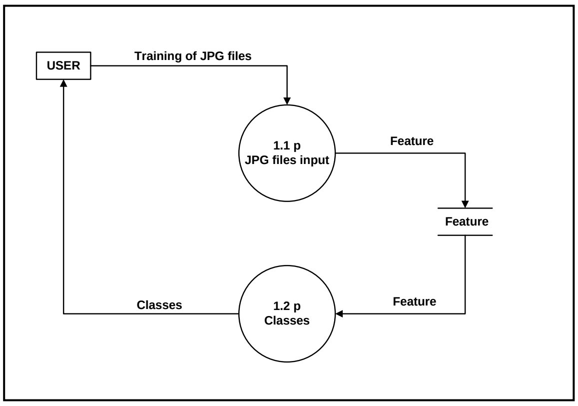

Figure 5: Primitive DFD of process 1.0 Class determination

The detailed processes of Process 2.0 is displayed in Figure 6 where Process 2.1 p accepts the JPG file input as testing data, proceed to extract its features, then hand it over to Process 2.2 p which performs NCC classification with already-stored features of training data from the Feature data store. The results are transferred to Process 2.3 p which determined the class for the test data and return the classification result for the test image to the external entity USER. Therefore all the conceptual overview of the system can be completely perceived through the series of DFD Diagram.

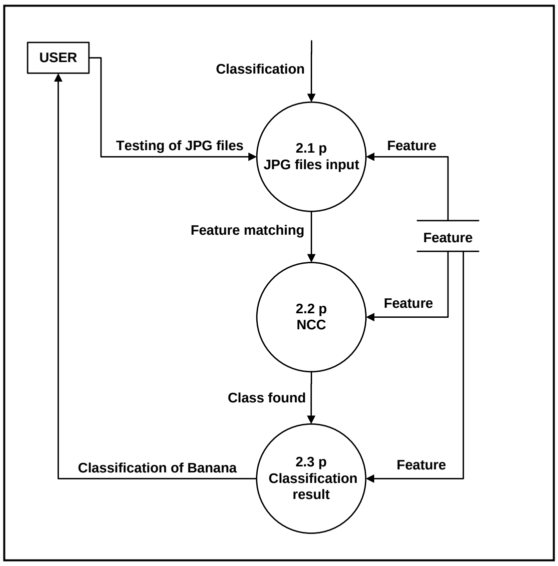

Figure 6: Primitive DFD of proses 2.0 Classification

Features from training images of all three types of banana are collected in the Feature data store and recalled to be classified against the incoming features from the test data. This research employs Nearest Centroid Classifier (NCC) to perform the classification process. In machine learning, a classification model that assigns the class label of a trained data set whose centroid is nearest to the data observed is the nearest centroid classifier or prototype classifier. Since the centroids indirectly represent the mean of each class, its distance from the test data can be computed to represent similarity of the data in question to the corresponding class members. Distance to the data is determined for all centroids and the smallest value is then selected. The observed data then assigned membership of the class with the nearest centroid [8].

NCC in this study employed Euclidean distance to measure the proximity of a test data to the centroids of each class, where the smallest distance belongs to the class where the test data will be assigned to. Euclidean distance d between two instances (p, q) is calculated using the following formula in Equation (3), where 1..n represents the number of features. In this research, they represent each H, S and V values.

Since we have four different classes, every test data undergone the distance calculation four times, each against centroids of green, almost ripe, ripe and overripe classes.

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} \quad (3)$$

p : test data

q : centroid of a class

$p_1..p_n$: features 1..n of the test data

$q_1..q_n$: features 1..n of the centroid

The distribution of training data used to form the centroids of four classes (green, almost ripe, ripe, overripe) are shown in Figure 7.

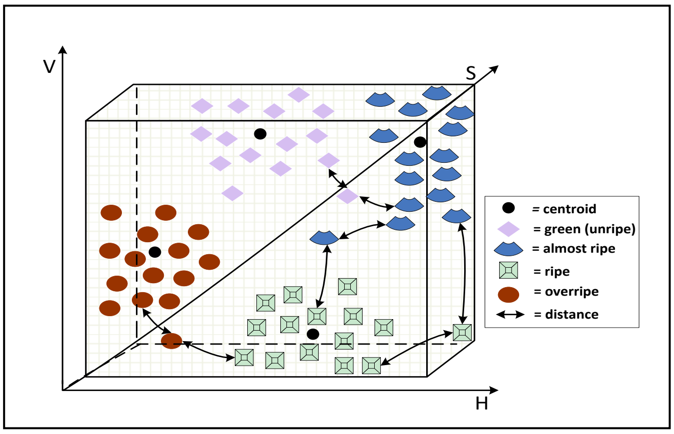

Figure 7: Data distribution in HSV colour space with NCC

The training database is formed using 60 image examples of the three kind of bananas divided into four classes (green, almost ripe, ripe and overripe). Range of features obtained for each classes are shown in Table 2.

Table 2: Class Ranges

	Green (Unripe)	Almost Ripe	Ripe	Overripe
Range of feature	2151.6	2638.0	2345.2	2046.4

4 Result and Discussion

Testing is performed by using 60 mixed images of *ambon lumut*, *kepok* and *raja* bananas: 15 green, 15 almost ripe, 15 ripe and 15 overripe, to be classified against the collected training data. Each kind of banana is not tested separately, so every image is tested against a database of mixed banana types. The classification results are presented as confusion matrix in Table 3. The left heading is the input test data of each class, while the top heading displays the classification results.

Table 3: Confusion Matrix

	Green (unripe)	Almost Ripe	Ripe	Overripe
Green (unripe)	14	1	-	-
Almost Ripe	-	12	-	3
Ripe	-	-	10	5
Overripe	-	-	7	8

From Table 3 it is clear that the green or unripe test data is the best to be classified with only one fault of 15 examples (93.33% accuracy), while the worst is the overripe which only 8 of 15 is classified correctly (53.33% accuracy). The second best is the almost ripe which the system correctly classified 12 out of 15, but with 3 of them classified as overripe. This might be caused by similar colour hue of labeled ‘overripe’ examples of another banana type. The ripe class itself is 66.67% correctly classified (10 of 15) with acceptable mistook class of overripe.

The green examples in test data is easily classified since all three kinds of banana agreed in colour during this stage of ripeness. As the ripening progressed, each kind is showing its own characteristic of ripe colour: *ambon lumut* moss green or pale bright green, *kepok* yellow and *raja* yellow with a very faint orange tone. These differences explain the confusion between almost ripe, ripe and overripe classes. The research thus shows the possibility of using HSV as a feature in indicating ripeness stage of banana, albeit it is not certain whether this feature is applicable to detect the ripeness of any kind of banana using a single classification device.

5 Conclusion

Experiments performed on this research to test whether HSV colour features are applicable to classify ripeness stage of ambon lumut, kepok and raja banana have the following conclusions:

1. When used as colour features, HSV values are able to distinguish four classes of ripeness stage (green, almost ripe, ripe and overripe) with ripe and overripe classes shared an overlapping borderline, so that one another may wrongly classified. This may be caused by the small difference in colour tone of ripe and overripe, only differentiated by the darkness of the peel.

2. The removal of the background affects the experiment result, which increase the ability to distinguish ripe and overripe classes.

3. Average classification rate is 73.33%.

Over the findings, we have the following suggestions to improve similar future researches:

1. Although replacing the background with white improves the success rate of the classification, we suggest in employing a segmentation method over the images, so that only parts of images containing banana is in consideration for the feature extraction.

2. Employs other classification methods such as k-Nearest Neighbour (k-NN) algorithm as a comparison to Nearest Centroid Classifier (NCC) for a possibility of improvement.

3. As the colour of a starting-to-overripe banana peels includes dark freckles during the process to be entirely brown, we suggest adding a texture feature to strengthen the classification between ripe and overripe bananas.

References

[1] R. Singh, R. Kaushikand and S. Gosewade, "Bananas as underutilized fruit having huge potential as raw materials for food and non-food processing industries: A brief review," The Pharma Innovation Journal, Vol. 7 (6): pp. 574–580, May 2018, https://www.thepharmajournal.com/archives/2018/vol7issue6/PartI/7-6-103-451.pdf, accessed 10 August 2020.

[2] H.T. Vu, C.J. Scarlett and Q.V. Voung, "Phenolic compounds within banana peel and their potential uses: A review," Journal of Functional Foods, Vol. 40: pp. 238–248, January 2018, https://www.sciencedirect.com/science/article/abs/pii/S1756464617306783, accessed 10 August 2020.

[3] L. Hapsari and D.A. Lestari, "Fruit Characteristic And Nutrient Values Of Four Indonesian Banana Cultivars (Musa Spp.) At Different Genomic Groups," Agrivita Journal of Agricultural Science, Vol. 38 (3): pp. 303–311, May 2016.

[4] A. Pereiran and M. Maraschin, "Banana (Musa spp) from peel to pulp: Ethnopharmacology, source of bioactive compounds and its relevance for human health," Journal of Ethnopharmacology, Vol. 160: pp. 149–163, November 2014, https://www.natural-knowhow.com/wp-content/uploads/2015/07/Banana-Musa-spp-from-peel-to-pulp-Ethnopharmacology-source-of-bioactive-compounds-and-its-relevance-for-human-health-article.pdf, accessed 19 August 2020.

[5] A.W.K. To, G. Paul and D. Liu, "Surface-Type Classification Using RGB-D," IEEE Transactions on Automation Science and Engineering, Vol. 11 (2): pp. 359–366, April 2014, https://ieeexplore.ieee.org/document/6675887, accessed 12 September 2020.

[6] C.D. Chang, S.S. Yu, H.H. Chen and C.S. Tsai, "HSV-based Color Texture Image Classification using Wavelet Transform and Motif Patterns," Journal of Computers, Vol. 20 (4): pp. 63–69, January 2010.

[7] D. John, "December Communication," RGB to HSV color conversion, April 2020, https://www.december.com/html/spec/colorhsltable.html, accessed 30 July 2020.

[8] E.N. Tamatjita and A.W. Mahastama, “Comparison of Music Genre Classification Using Nearest Centroid Classifier and k-Nearest Neighbours,” 2016 International Conference on Information Management and Technology (ICIMTech): pp. 118–123, May 2017, https://ieeexplore.ieee.org/document/7930314, accessed 30 July 2020.